A New Kind of Biological

Advanced simulations are revealing the exact mechanisms of key biological processes.

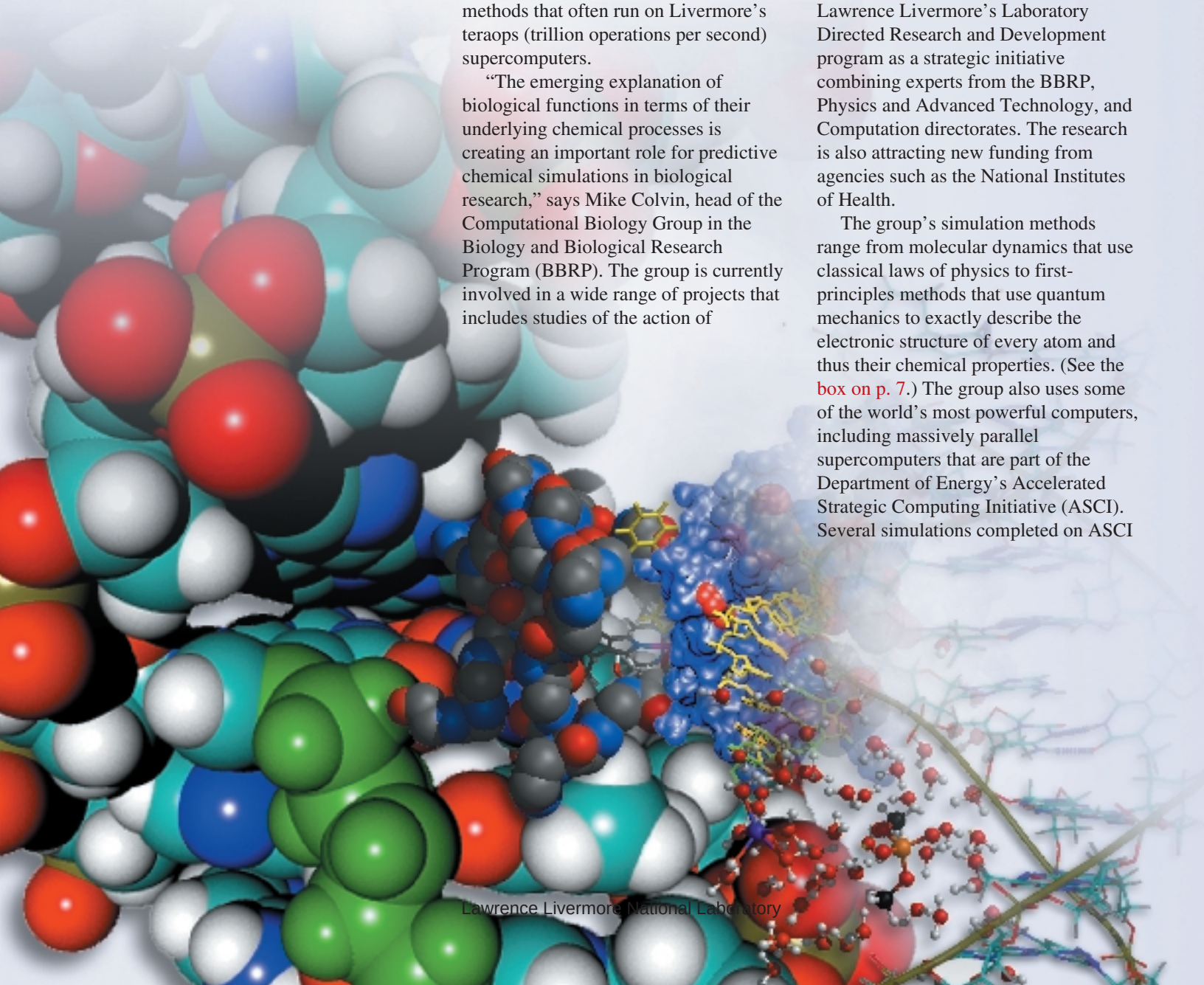
A team of Lawrence Livermore researchers is developing a new approach to biological research by linking advanced simulations with laboratory experiments to explain biological phenomena at an unprecedented level of detail. In addition to promoting close collaborations with other biological scientists at the Laboratory, the research involves newly developed simulation methods that often run on Livermore's teraops (trillion operations per second) supercomputers.

"The emerging explanation of biological functions in terms of their underlying chemical processes is creating an important role for predictive chemical simulations in biological research," says Mike Colvin, head of the Computational Biology Group in the Biology and Biological Research Program (BBRP). The group is currently involved in a wide range of projects that includes studies of the action of anticancer drugs, the DNA-binding properties of mutagens in food, the mechanisms of DNA repair enzymes, and the biophysics of DNA base pairing.

All of the projects strongly tie modelers to experimental researchers in BBRP because, says Colvin, "There is a growing consensus that the integration of computation and experiment will accelerate progress in biology."

The work is funded in part by Lawrence Livermore's Laboratory Directed Research and Development program as a strategic initiative combining experts from the BBRP, Physics and Advanced Technology, and Computation directorates. The research is also attracting new funding from agencies such as the National Institutes of Health.

The group's simulation methods range from molecular dynamics that use classical laws of physics to first-principles methods that use quantum mechanics to exactly describe the electronic structure of every atom and thus their chemical properties. (See the box on p. 7.) The group also uses some of the world's most powerful computers, including massively parallel supercomputers that are part of the Department of Energy's Accelerated Strategic Computing Initiative (ASCI). Several simulations completed on ASCI



Lawrence Livermore National Laboratory

Research

computers have established new standards for size and accuracy in the chemical modeling of biological processes.

DNA Lesions Need Repair

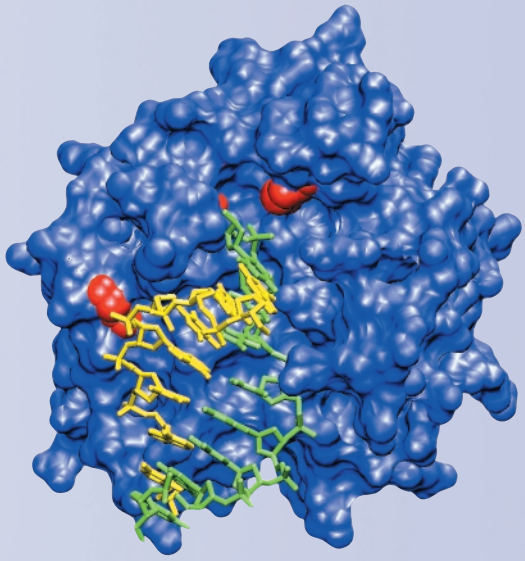
Much of the computational biology work focuses on the dynamics and structure of DNA and repair mechanisms for specific forms of DNA damage. Human genes are made up of long, double-stranded DNA molecules that contain the instructions for building the proteins that make up the machinery of every cell. Understanding the role of DNA repair and other processes that protect cells from radiation and chemical insults has been a long-term interest at Lawrence Livermore.

One simulation project has shown how human repair enzymes recognize a common form of DNA damage called an abasic lesion. These lesions occur when a stretch of DNA loses one of its constituent bases, leaving a gap in the chain. Such lesions arise spontaneously more than 10,000 times per day in every cell. They can also be caused by exposure to pesticides, food mutagens, and ionizing radiation from the sun. If unrepaired, the damage can lead to disease, notably cancer, through a mutation in the genetic code.

Cells have a set of repair enzymes (proteins) that scan DNA looking for damage such as abasic lesions. In humans, the major enzyme responsible for repairing abasic DNA is the endonuclease Ape1. Livermore scientists have been addressing the question of how repair proteins such as Ape1 recognize and bind specifically to damaged DNA.

"We want to know what changes to the DNA shape caused by the lesions are recognized by the repair protein," says Computational Biology Group member Daniel Barsky. Until his simulations were completed last year, scientists had suggested a variety of possibilities.

Using a structurally related protein–DNA complex as a template, Barsky first built a model of DNA bound to Ape1 to guide BBRP biologist David Wilson III and colleagues in determining which amino acids of Ape1 might be most important for its activity. Wilson also synthesized different forms of DNA to see which, if any, would be recognized by Ape1. Surprisingly, all of the altered forms attracted Ape1 to a significant degree. Spurred by Wilson's findings, Barsky used a cluster of advanced scientific workstations in the Livermore Open Computing Facility to complete the first full molecular dynamics simulations of the damaged DNA.



A model of Ape1, a DNA-repair protein, binding to a strand of DNA. DNA strands are in yellow and green, the protein is in blue, and ultraviolet-absorbing amino acids are in red.

The simulations focused on a stretch of DNA with an abasic lesion surrounded by thousands of water molecules. Although the computations simulated a time span of only 2 billionths of a second, they required the equivalent of several months of processing time on a fast single-processor computer.

Shape Changes Flag Enzyme

The results showed that the abasic site does not form a permanent hole or gap in the DNA, as some researchers had postulated. Instead, the missing base causes changes to the internal motions of the DNA—changes that are thought to be important for damage recognition by the repair enzyme.

For example, at various intervals in Barsky's simulations of damaged DNA, a thymine base unpaired with adenine and paired instead with the cytosine base that was opposite the abasic site, where a guanine would normally be. This transient thymine–cytosine base-pair mismatch had not been previously observed. Another clearly visible change was that the sugar molecule, formerly attached to the missing guanine, flipped out of the chain.

The results indicated that the abasic DNA chain has a great deal more flexibility and bending than normal. This unnatural flexibility apparently "flags" the repair protein, says Barsky. Further proof for this concept was supplied last year when x-ray crystallography studies of abasic DNA bound to Ape1 showed a kink in the DNA at the abasic site.

The results from the simulation are being used to determine how specific

differences in the Ape1 protein that have been found in a portion of the human population affect the DNA repair capacity of those individuals. Such knowledge will help researchers predict which people are at greater risk of developing disease from environmental exposures that induce the formation of abasic lesions.

Barsky has also been studying the base pairings in parallel-stranded DNA, a novel form of DNA in which both strands are oriented in the same direction instead of aligning in opposite directions. His quantum chemical calculations contradict previous theories of how the guanine–cytosine base pairing might occur in parallel DNA. The results predict that the greatest stability occurs through an orientational "wobble" in which guanine and cytosine form only two hydrogen bonds instead of the three bonds they form in normal DNA.

Zooming In on Key Ions

Barsky's DNA simulations are magnified a millionfold in simulations done by colleagues Eric Schwegler and Felice Lightstone. Their focus is on the chemical reactions involved in phosphate hydrolysis, an essential part of DNA repair. To study this reaction, they first slice a strand of DNA in two to see how other enzymes repair the damage.

Lightstone's first-principles molecular dynamics simulations use 65 water molecules, one dimethyl phosphate (the simplest repeating structure comprising the DNA backbone), and one magnesium ion in a cubic "box." The model includes a magnesium ion because many DNA-cutting enzymes such as Ape1 require the ion to catalyze the DNA cleavage. In fact, a high concentration of magnesium ions alone can catalyze the cleavage.

The simulations track the movement of every atom and its cloud of electrons. Of particular interest is what happens to the magnesium ion. It attracts six water molecules and momentarily adheres to the dimethyl phosphate, thus making the cleavage reaction possible.

The simulations are run on the ASCI Blue supercomputer using Lawrence Livermore software adapted for multiprocessor machines. Despite the enormous computational power of the computer, Lightstone can only simulate one trillionth of a second per month. "As a result," says Lightstone, "we have to be selective in what we simulate."

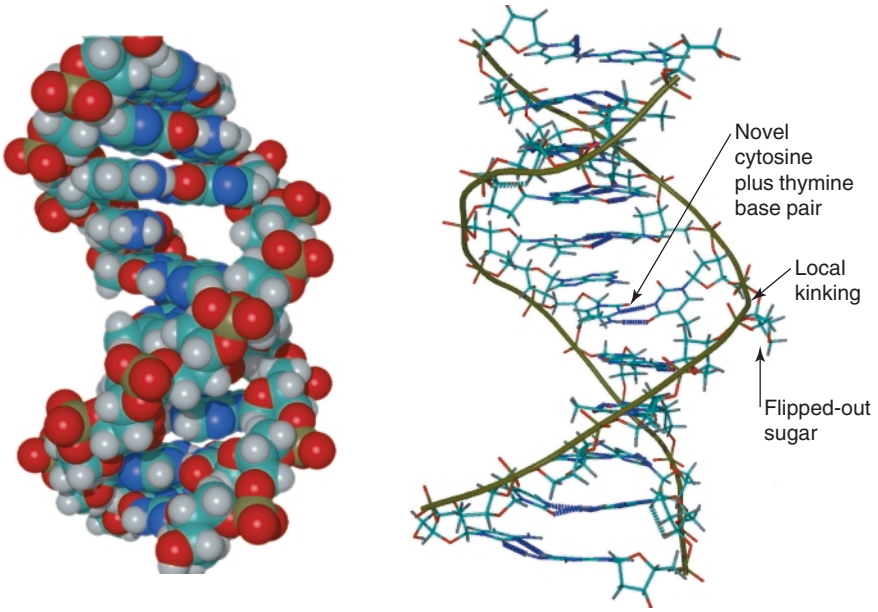
The payoff is new understanding of the pathways of reactions involved in DNA repair. "Experimentalists can't tell you all the steps involved in a biological process," Lightstone says. "Seeing all the steps gives us ideas about how we might modify the reaction or even control the enzyme's actions."

Selective Docking to Enzymes

Lightstone is also working on so-called computational docking to help identify small molecules called ligands. Ligands can uniquely bind to selected sites on proteins, including DNA-repair enzymes and deadly neurotoxins such as tetanus and its relative, botulinum. The identified ligands would be used in sensors to indicate the presence of neurotoxins. The effort is being pursued in conjunction with BBRP researcher Rod Balhorn, Livermore's Chemical and Biological Nonproliferation Program, and computational experts at Sandia National Laboratories. (See *S&TR*, April 1999, pp. 4–9.)

Proof-of-principle computational docking calculations were first used to screen the Available Chemicals Directory—a listing of some 250,000 purchasable compounds for which the three-dimensional structure is known—for chemicals that would bind to the tetanus protein. A cluster of scientific workstations using molecular mechanics techniques required only 10 seconds to analyze the shape of each candidate molecule and determine the extent, if any, to which it could bind to a small depression in the tetanus protein's surface.

After three days of calculations, the simulation had ranked all of the compounds in the chemical library, out



The model at left shows a DNA molecule missing one of its bases. The simulation at right reveals its kinked, unnatural shape that "flags" the repair protein Ape1.

Using a Virtual Microscope

For Lawrence Livermore computational biologist Mike Colvin, the most difficult aspect of designing and running a biological simulation is fully understanding the biological problem at hand and reducing it to one or more key chemical reactions. An example is determining why one molecule is vastly more toxic to a cell than another chemically similar molecule.

"The first step in our job is to dig through the literature and talk with the experimental collaborators to identify the essential reactions that we need to simulate in order to discover the differences in toxicity," says Colvin. Such simulations reveal the exact physical mechanisms and energetics of the process.

Colleague Daniel Barsky calls advanced biochemical simulations a "virtual microscope." This microscope, he explains, combines powerful computers and software programs that link the laws of physics and chemistry to structures of biological molecules that have been determined by x-ray diffraction and other experimental methods.

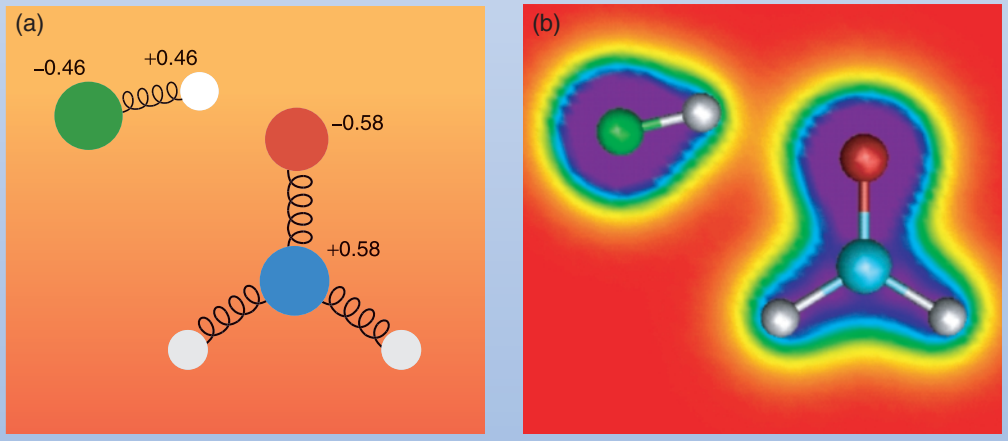
He cautions, however, that one danger with the microscope analogy is that scientists new to such techniques may be tempted to simulate everything without carefully identifying the biological question being addressed. "We have found that the simulations must be very closely tied to experiments. Without having specific things to test and look for, your simulations may produce a mountain of data that do not reveal much."

The biochemical simulations principally take two forms: molecular dynamics and first-principles quantum chemistry. Molecular dynamics is used for DNA and protein studies involving tens of thousands of atoms at a time. These simulations depict atoms like balls interconnected with springs to approximate their motions and their interactions with other atoms.

Simulations based on first-principles quantum chemistry accurately predict the chemical properties of atoms and molecules. The technique uses quantum mechanics to determine the distribution of electrons around each atom. From this electron distribution, any chemical property can be determined, including the structures and energies of molecules.

Teraops-scale supercomputers have made it possible to do so-called first-principles molecular dynamics in which the motion of molecules is simulated using accurate quantum chemical interactions. These methods have been developed by Lawrence Livermore physicists Francois Gygi at the Center for Applied Scientific Computing and Eric Schwegler and Giulia Galli in the Physics and Advanced Technologies Directorate. The new techniques are being applied to selected biochemical problems, such as how DNA is cleaved by repair enzymes and how an anticancer drug is activated by the body.

Frequently, different simulation methods are used together to solve a single problem. For example, a researcher can do molecular dynamics on DNA structure, then pull out a small area of interest and do first-principles simulations for understanding the exact electronic forces at play.



Examples of two modeling methods applied to the interaction of hydrogen fluoride and formaldehyde. (a) Molecular dynamics models depict atoms like balls interconnected with springs to simulate their motions. (b) First-principles simulations use quantum mechanics to predict the chemical properties of atoms and molecules.

of which Lightstone compiled a list of 11 candidates for laboratory tests. BBRP experimentalists found that 5 of the 11 molecules successfully bound to the tetanus protein.

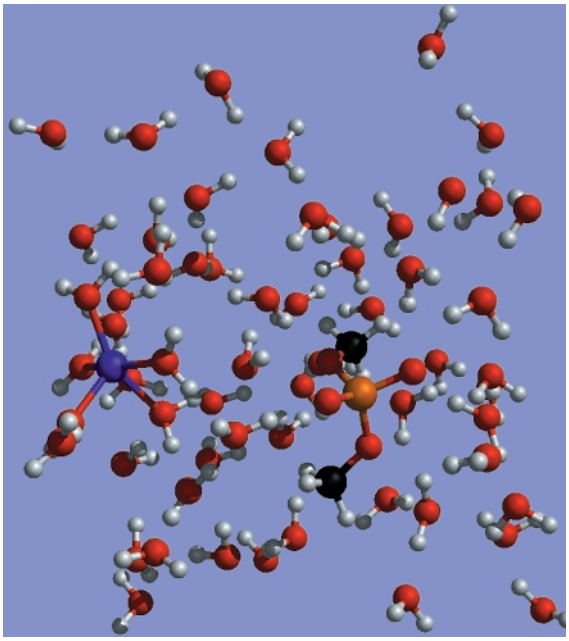
Lightstone points to the efficiency of using methods such as computational docking as a screening tool. "Experimentally testing 250,000 compounds would take years of work," she says. She notes that the docking simulations are done at a coarse level of accuracy. "If you did them at first-principles level, we would get bogged down in long computer times. Our purpose is to screen lots of compounds very quickly."

Lightstone moved on to finding candidate ligands for binding to botulinum toxin, which is viewed as a more dangerous threat. This time, she applied a second computational step, flexible docking, to the top 2,000 compounds identified in the first step. Flexible docking, which requires two additional weeks of computer time, rotates molecular bonds to find the optimum shape of a molecule for binding to a selected protein. The top-scoring ligands are being tested for their affinity to the botulinum protein.

To increase the sensitivity and selectivity of a portable sensor, Lightstone is searching for an additional ligand for a second, nearby site on the botulinum protein. Together, the two ligands could be used to detect the toxin at low concentrations.

Mutagens Get Activated

The longest running collaborative project in the Computational Biology Group is studying the function of mutagenic chemicals called heterocyclic amines. These compounds are formed in the cooking of several foods and may be a risk factor associated with cancer in the human digestive tract. As with most substances that damage DNA and cause cancer, the food mutagens must be activated by metabolic reactions the body uses to break down chemicals. Once activated, the mutagens bind to DNA and can interfere with the accurate duplication of the genetic code, thereby leading to mutations and eventually cancer.



This simulation focuses on magnesium, which is required by many DNA repair enzymes. The simulation includes many water molecules and one dimethyl phosphate molecule (the simplest repeating component of DNA). Oxygen is red, hydrogen is white, carbon is black, magnesium is purple, and phosphate is orange.

For the past four years, the group has been applying simulations to help experimentalists identify a subset of heterocyclic amines called 2-aminoimidazole-azarene (AIA) compounds. The group is looking at several properties of these mutagens, including the metabolic steps that actuate the compounds, the initial attachment site of these mutagenic compounds on DNA, how binding to the mutagens depends on the DNA sequence of bases, and the effect of different cooking processes on mutagen formation.

"A paradox is that all two dozen molecules in the AIA family are chemically similar, and yet there is a 10-millionfold range in mutagenicity," says Colvin. Identifying the factors that vary the potency will help scientists better predict the human health risks associated with exposure to food mutagens.

The most recent simulations involve the action of cytochrome P450, the enzyme responsible for first activating the mutagens. Because the structure of the human form of P450 is unknown, the group built a computer model using related enzymes whose structure is known. The simulations are used to dock different AIA compounds into the enzyme's active site to determine if the more potent species make a better fit and are thus more likely to become activated.

The group is also simulating the interaction between P450 and members of the bioflavanoid family, which inhibit food mutagens. About the same size and shape as AIA compounds, bioflavanoids are found in fruits and vegetables. One hypothesis is that the bioflavanoids lower the incidence of food mutagens by competing for the same activation site on P450 as the AIA compounds.

A Virtual Frypan

A new effort in the Computational Biology Group is to develop an accurate computer model of cooking hamburger patties and other meats. The goal, says student researcher Ngoc Tran, is to successfully simulate what BBRP investigators do on a hot stove to study how cooking methods affect food mutagen production. Thus, virtual cooking would help reduce the number of experiments done in the kitchen by identifying the key experimental measurements that must be made.

Developing a simulation of the cooking process and the formation of mutagens is not straightforward because it must reflect such factors as the fat and moisture content of the meat, the frying temperature, the heat conductivity of the pan and the meat, and the manner of cooking. For example, BBRP researchers discovered last year that flipping a hamburger frequently during frying reduces the number of mutagens. (See *S&TR*, January/February 2001, p. 2.)

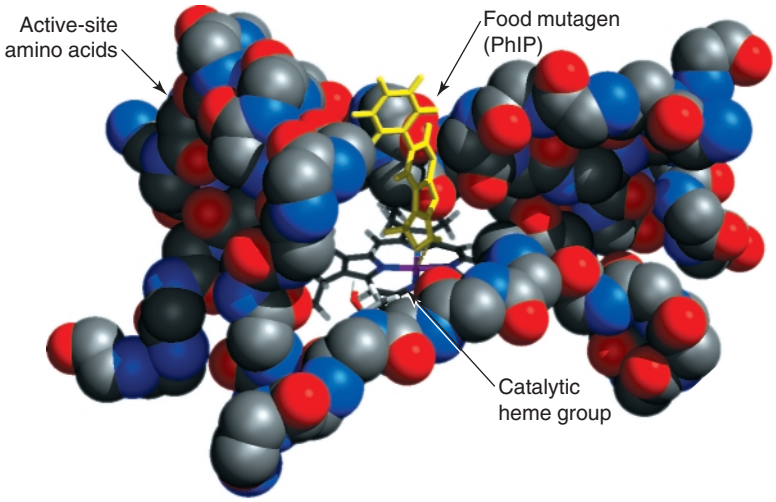
Tran spent last summer working with Mark Knize and Cyndy Salmon accumulating raw data for the simulations by cooking a couple dozen hamburgers. She measured the temperature at different depths of the patties as well as the corresponding concentrations of AIA compounds. She found that the first millimeter contained 50 percent of the food mutagens, the second millimeter contained 25 percent, and the third contained 10 to 15 percent. The simulations accurately reproduced the temperature profiles measured while cooking beef patties and correctly predicted how the concentration of mutagens varied at different meat depths.

The current goal is to refine the model so that it accurately reflects every aspect of cooking. Simulations will then be used to determine the sensitivity of mutagen formation to such parameters as fat content and pan temperature. The ultimate goal is to design new cooking procedures that minimize the formation of mutagens.

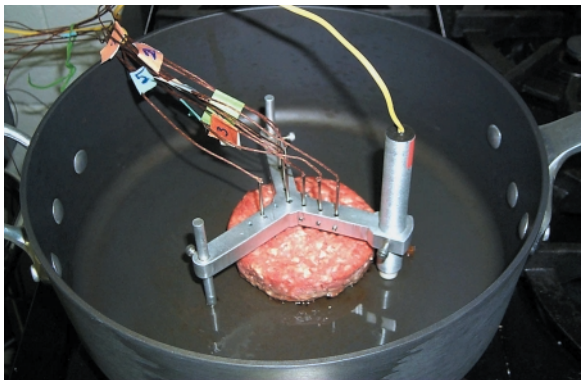
Optimizing Anticancer Drugs

Simulations are also helping scientists to understand the functioning of one of the oldest families of anticancer drugs, the phosphoramidic mustards. The group includes the widely used drugs cyclophosphamide and ifosfamide. These drugs are closely related to the poisonous mustard gas used in World War I. Doctors noticed at the time that the gas killed rapidly dividing body cells and reasoned that a derivative might work on cancer cells because they continually divide.

Despite being used for more than 40 years, several important questions about the drugs' biological activity remain unanswered. The key to the drugs' therapeutic activity—and toxic side effects—are the activation steps they undergo in the body before they bind to DNA.



This simulation reveals the binding mechanism between a food mutagen and cytochrome P450, the enzyme that catalyzes the initial activation step for this mutagen.



A fully instrumented hamburger patty is fried to determine its temperature as a function of depth as well as the corresponding concentrations of food mutagens. The data are used to develop computer simulations of the cooking process and to predict the formation of mutagens.

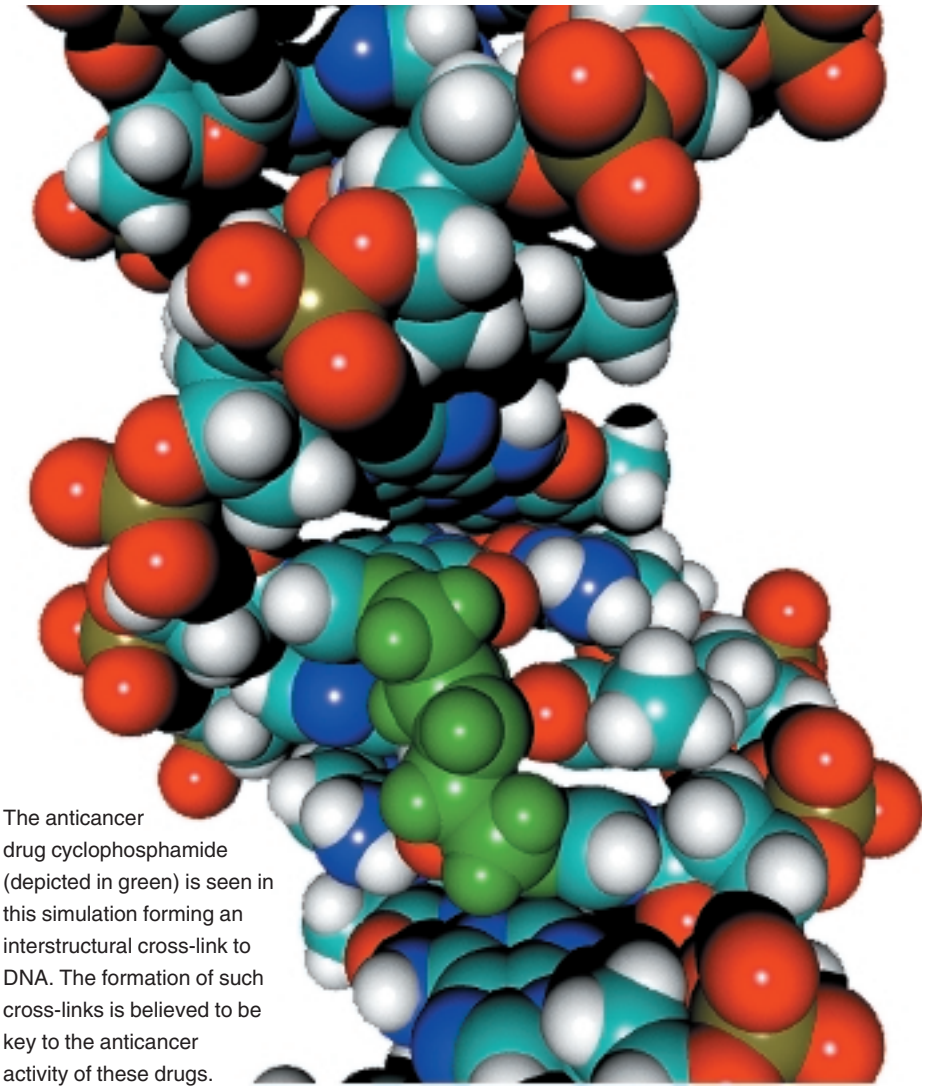
Colvin has been collaborating with scientists at Duke University's Comprehensive Cancer Center to understand how the activation reactions affect the drugs' potency. Fortunately, the drugs are small enough to be modeled using first-principles methods. Simulations of most of the steps in the activation pathway have helped to explain several unexpected properties of these drugs and are suggesting improved versions of the standard mustard drugs.

Mustard drugs are known to kill cancer cells by forming cross-links between the two strands of a cell's DNA. These cross-links are particularly difficult for a cell to eliminate; just a few cross-links kill a cancer cell. Colvin and Dat Nguyen, a graduate student researcher from the University of California at Davis, are simulating how cyclophosphamide forms cross-links. The goal is to understand how the drug's molecular structure could be changed to make it more effective in forming cross-links.

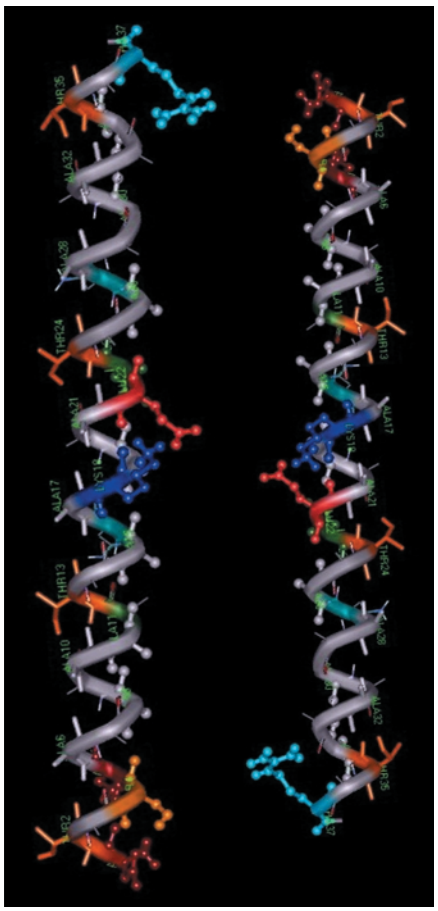
Researchers at Duke University and the University of Chicago will synthesize the best molecular candidates identified in the simulations and test their cross-linking capacity on DNA. The most promising molecules will then be tested for their effectiveness in killing cancer cells.

Antifreeze Needed Here

The Computational Biology Group is also collaborating with Nguyen and several UC Davis professors to



The anticancer drug cyclophosphamide (depicted in green) is seen in this simulation forming an interstructural cross-link to DNA. The formation of such cross-links is believed to be key to the anticancer activity of these drugs.



Molecular dynamics simulations indicate how antifreeze proteins in the winter flounder may prevent ice formation by forming several stable structures, like this one, that act as an insulating blanket.

examine a phenomenon that has puzzled biologists for more than four decades: the ability of Antarctic fish to survive sea temperatures well below freezing. The fish contain a variety of so-called antifreeze proteins in their bloodstream that inhibit the growth of ice crystals in their bodies. Similar proteins have also been identified in some insects and plants that can withstand freezing temperatures.

The exact mechanisms of how they function have been a mystery, but molecular dynamics simulations by Nguyen provide new clues. The simulations tested the hypothesis that these proteins depress the freezing temperature by binding to ice crystals and acting like an insulating blanket. The results showed that pairs of antifreeze proteins can form several stable structures. The proteins may be able to absorb and store heat by undergoing transitions between these structures.

Simulations' Value Recognized

Colvin says that Lawrence Livermore's new capabilities in computational biology are being recognized by the greater scientific research community in the form of invited talks, requests for review articles and textbook chapters, and new collaborations by colleagues at universities and companies.

"As computational modeling is successfully applied to more biological problems," Colvin says, "it is clear that simulation will have a growing role in the training and research of biological scientists." There is also little doubt that advanced simulations will continue to change the nature of biological research.

—Arnie Heller

Key Words: abasic lesions, Accelerated Strategic Computing Initiative (ASCI), Ape1, bioflavanoids, botulinum, Center for Applied Scientific Computing, computational biochemistry, cyclophosphamide, DNA, first-principles molecular dynamics, food mutagens, molecular dynamics, quantum mechanics, tetanus.

For further information contact Mike Colvin (925) 423-9177 (colvin2@llnl.gov).

About the Scientist

MIKE COLVIN is leader of the Computational Biology Group in Livermore's Biology and Biotechnology Research Program (BBRP). He received B.S. degrees in chemistry and humanities from the Massachusetts Institute of Technology and a Ph.D. in chemistry from the University of California at Berkeley. He joined the Laboratory in 1986 as a postdoctoral fellow in the Institute of Scientific Computing Research and then became a staff physicist in O Division, where he concentrated on designing biologically inspired algorithms for faint object detection. In 1990, Colvin transferred to the Center for Computational Engineering at Sandia National Laboratories, California, to develop quantum chemical methods for massively parallel computers. In 1997, he returned to Livermore and joined BBRP. His research there has focused on using advanced computer simulations to study biological phenomena.